

Stochastic Relational Presheaves and Dynamic Logic for Contextuality

Kohei Kishida*

Department of Computer Science
University of Oxford
Oxford, United Kingdom
kohei.kishida@cs.ox.ac.uk

Presheaf models [7, 27, etc.] provide a formulation of labelled transition systems that is useful for, among other things, modelling concurrent computation. This paper aims to extend such models further to represent stochastic dynamics such as shown in quantum systems. After reviewing what presheaf models represent and what certain operations on them mean in terms of notions such as internal and external choices, composition of systems, and so on, I will show how to extend those models and ideas by combining them with ideas from other category-theoretic approaches to relational models [15] and to stochastic processes [11, 3, 17, etc.]. It turns out that my extension yields a transitional formulation of sheaf-theoretic structures that Abramsky and Brandenburger [1] proposed to characterize non-locality and contextuality. An alternative characterization of contextuality will then be given in terms of a dynamic modal logic of the models I put forward.

1 Introduction

The goal of this paper is to devise a formalism of semantic structure for dynamic logic that is suitable for expressing stochastic dynamics such as shown in quantum systems. Essential features of stochastic dynamics I aim to capture include

- the distinction and interaction between *internal* and *external choices*, that is, non-deterministic branchings that are made within a system and that are made by external agents or experimenters;
- the distinction and interaction between what is *globally* the case in an entire system and what is *locally* the case in a subsystem.

In particular, the resulting semantics and logic shall be general enough to accommodate both the presence and the absence of (typically quantum) *non-locality* and *contextuality*, but at the same time expressive enough to provide logical characterization for non-locality and contextuality.

I achieve my goal by integrating three frameworks of categorical approaches that have been proposed to modelling non-deterministic and stochastic processes. Firstly, my formalism will be based on

- (i) Presheaves as labelled transition models for concurrency (Winskel *et al.* [7, 27], etc.). I show how the presheaf structure can be used to capture notions that are essential to my goal, such as internal and external choices, composition of multipartite systems, and so on.

*Kishida's research has been supported by the grant FA9550-12-1-0136 of the U.S. AFOSR. He is grateful to the audience at QPL 2014 as well as to the anonymous referees for their insightful comments and suggestions, which improved this paper. Grateful acknowledgment also goes to the audience at the 2013 Workshop on Information and Processes, at which an earlier version of this paper was presented, and in particular to Prof. Michael Mislove and Tulane University, who hosted and funded the workshop.

Then I extend this setting in two aspects, by admitting non-trees and by adding probabilities. I attain these extensions by integrating the following ideas into my formalism.

- (ii) Kripke relational semantics in terms of Kripke frames as functors from labels to the category **Rel** of sets and relations (Hermida [15]). Integrating this idea with the presheaf framework admits presheaf-like models as transition systems of non-tree forms. I will also lay out motivation for admitting non-trees. (One mode of this integration has already been given in Sobociński [25]; yet the mode of integration I propose in this paper is different and not equivalent.)
- (iii) The category of stochastic maps, or equivalently the Kleisli category of the distribution monad (the idea goes back at least to Lawvere [18]; it is also studied recently by Fritz [11], Baez *et al.* [3], Fong [9], etc., in the former formulation, by Jacobs [17], etc., in the latter formulation). How to add probabilities to presheaf models is a question posed in the concluding part of Varacca [26]; I answer this question by using structures closely related, though not equivalent, to stochastic maps.

These extensions give semantic structures on which I define a dynamic and probabilistic logic.

To demonstrate that the resulting semantics and logic achieve the goal mentioned above, I will show how they capture non-locality and contextuality. In particular, the semantics gives an alternative, transitional formulation to a sheaf-theoretic approach to non-locality and contextuality (Abramsky and Brandenburger [1], etc.). This approach provides a sheaf-theoretic expression for, among other things, measurement scenarios in quantum mechanics, and characterizes non-locality and contextuality found in such scenarios in terms of non-existence of global sections. The transitional formulation I give to this approach leads to an alternative, dynamic-logical characterization of non-locality and contextuality.

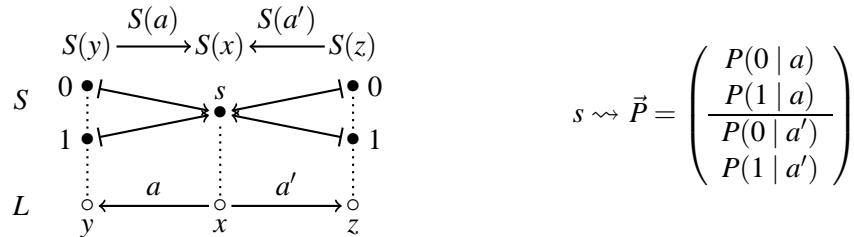
2 Presheaf Models for Measurements

This section reviews presheaves over trees as labelled transition systems (see [7, 27]). Rather than giving new definitions or theorems, this section is concerned with conceptually laying out how to use the familiar notions of presheaf and fibration to represent features of non-deterministic processes that are essential to the goal of this paper.

2.1 Trees and Presheaves of Non-Deterministic Choices

Here I lay out the key idea of how to use a presheaf over a tree as a labelled transition system, or LTS for short, in a manner suitable for representing different kinds of non-determinacy in stochastic processes.

As in the standard terminology, by a “measurement scenario of (n, k, ℓ) -type” let us mean a Bell-type scenario of (typically quantum) measurements that involves n parts (or experimenters), each of which (or whom) chooses one from k measurements, each of which has ℓ outcomes. For instance, in a $(1, 2, 2)$ scenario, Alice chooses one from two measurements, a and a' , each of which has two outcomes, 0 and 1. This simple scenario can be represented by the following tree L and presheaf S over L .



The binary branching in L represents the choice Alice makes outside the system, choosing from two measurements a and a' . Then regard S as a transition system, reading “ \mapsto ” backward as transition “ \leftarrow ”; each such edge of transition in S is labelled with an edge in L —for instance, those in $S(a)$ above are labelled with a , representing possible outcomes the system has for Alice’s choice of a . So the binary branching in $S(a)$ represents the system having two outcomes for measurement a . One of our objectives is to assign probabilities to such branchings, so that, in the picture above for instance, the state s can be (at least partially) specified by the vector of probabilities \vec{P} to the right of the picture above.

Note that the representation just given involves two kinds of choice. Put in general terms, when we describe a system and agents external to the system,

- The agents may be able to choose from different ways to interfere or interact with the system. We call these choices *external choices*, and represent them with branching in the base tree.
- The system may behave by itself non-deterministically—sometimes in response to external choices, but sometimes simply as time passes—with several possible outcomes. We call these choices *internal choices*, and represent them with branching in function components of the presheaf.

In short, external choice resides in the base tree L ; internal choice resides in (function components of) the presheaf S . This is the slogan for our use of presheaves S over trees L as L -labelled transition systems.

In fact, not just the distinction between internal and external choices, the presheaf structure also gives us several useful ways to control descriptions of these choices—for instance, to shift the boundary between the internal and external. We will see this in subsection 2.2. Before doing so, it is useful to observe that the presheaves over a tree are equivalent to the *fibrations* over the tree (which should be quite obvious from the picture above). Let us recall

Definition 1. A bundle (i.e., monotone map) $\pi : S \rightarrow L$ of posets is called a *fibration* (over L) if, whenever $x \leq_L \pi(t)$, there is a unique $s \in \pi^{-1}(x)$ such that $s \leq_S t$. Write **Fib** for the category of posets and fibrations.

We should note that if $\pi : S \rightarrow L$ is a fibration and L is a tree then S is also a tree. Then it is easy to show the following (we provide a proof rather as a review of notation).

Fact 1. $\mathbf{Sets}^{L^{\text{op}}} \simeq \mathbf{Fib}/L$ for any poset L .

Proof. A presheaf $S : L^{\text{op}} \rightarrow \mathbf{Sets}$ yields a fibration with the projection $\pi : S \rightarrow L$ from the dependent sum $S := \sum_{x \in L} S(x) = \{ (x, s) \mid x \in L \text{ and } s \in S(x) \}$ to L and the order \leq_S on S such that $(x, s) \leq_S (y, t)$ iff $x \leq_L y$ and $s = S(x, y)(t)$. A fibration $\pi : S \rightarrow L$ yields a presheaf $S : L^{\text{op}} \rightarrow \mathbf{Sets}$ by letting $S(x) = \pi^{-1}(x)$ for $x \in L$ and, whenever $x \leq_L y$, defining $S(x, y) : \pi^{-1}(y) \rightarrow \pi^{-1}(x)$ so that $S(x, y)(t)$ for $t \in \pi^{-1}(y)$ is the unique $s \in \pi^{-1}(x)$ such that $s \leq_S t$.

Given presheaves $S, T : L^{\text{op}} \rightarrow \mathbf{Sets}$ and corresponding fibrations $\pi_S : S \rightarrow L$, $\pi_T : T \rightarrow L$, the natural transformations from S to T are just the monotone maps $f : S \rightarrow T$ over L (meaning $\pi_T \circ f = \pi_S$), but any such monotone map f can easily be shown to be a fibration. \square

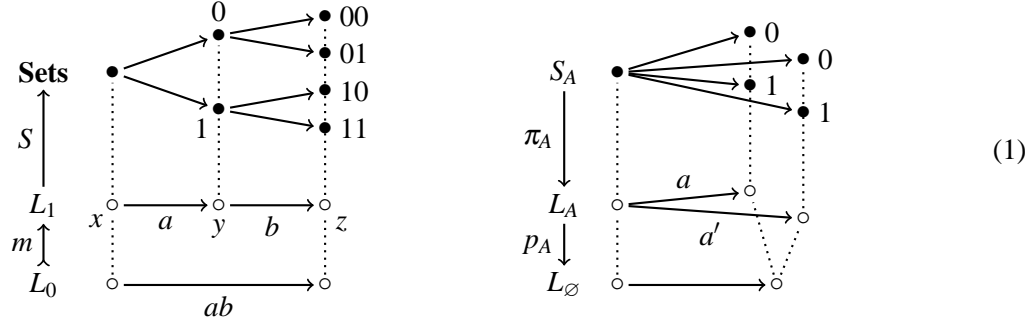
We will invoke this presheaf-fibration equivalence extensively in the rest of this paper.

2.2 Controlling System Descriptions

Given presheaf-fibration descriptions of non-deterministic processes with internal and external choices, we can take further advantage of operations on the presheaf-fibration structure to control the descriptions.

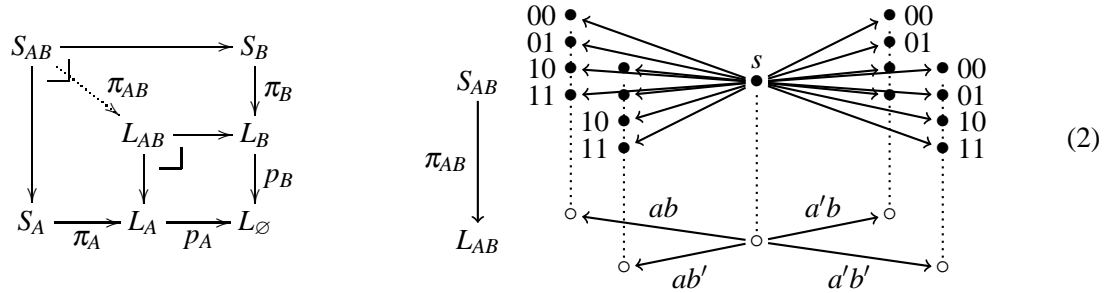
A family of operations that will later prove useful is done by change of base. One such operation is to precompose a given presheaf $S : L_1^{\text{op}} \rightarrow \mathbf{Sets}$ with an embedding $m : L_0 \hookrightarrow L_1$, obtaining a new presheaf $S \circ m^{\text{op}} : L_0^{\text{op}} \rightarrow \mathbf{Sets}$. Since some points, or “stages”, of L_1 are “omitted” in L_0 , the precomposition

makes the model “forget” what takes place at these omitted stages. For instance, take $m : L_0 \rightarrow L_1$ as on the left of (1) below, and let a and b represent measurements by Alice and by Bob. Then a presheaf $S : L_1^{\text{op}} \rightarrow \mathbf{Sets}$ carries information as to the original states (in $S(x)$), the possible outcomes of a (in $S(y)$), and then the possible further outcomes of b (in $S(z)$). In contrast, the presheaf $S \circ m^{\text{op}} : L_0^{\text{op}} \rightarrow \mathbf{Sets}$ carries the same information as to the original states (in $S(x)$) and the outcomes of both measurements (in $S(z)$), but it has no information as to the process in between (or, indeed, even as to whether a is performed before, after, or at the same time as b).



Another is to compose fibrations $\pi : S \rightarrow L_0$ and $p : L_0 \rightarrow L_1$, obtaining a new fibration $p \circ \pi : S \rightarrow L_1$. In π , branchings in L_0 represent external choices, but some of them are internal choices in p ; so the composition “internalizes” these external choices. Take π_A and p_A as on the right of (1) above. π_A describes Alice as an agent external to a system who externally chooses from measurements a and a' . On the other hand, $p_A \circ \pi_A$ describes a bigger system encompassing Alice—so that we simply watch the bigger system internally choose from the four outcomes, “Alice performs a and gets outcome 0”, etc.

In fact, such composition of fibrations can be used to compose descriptions of several systems into a description of a multipartite system. The fibration π_A in the picture above describes a $(1, 2, 2)$ -scenario for Alice. Take an isomorphic $\pi_B : S_B \rightarrow L_B$ to describe a $(1, 2, 2)$ -scenario for Bob. Then a fibration $\pi_{AB} : S_{AB} \rightarrow L_{AB}$ for the composed $(2, 2, 2)$ -scenario is obtained as follows:



That is, $\pi_{AB} = \pi_A \times_{L_\emptyset} \pi_B : S_A \times_{L_\emptyset} S_B \rightarrow L_A \times_{L_\emptyset} L_B$. Put more conceptually, we use L_\emptyset as a clock for synchronizing events in Alice’s scenario and ones in Bob’s, and then take simultaneous pairs of events from Alice’s and Bob’s scenarios. We should note that the pair of projections from S_{AB} and L_{AB} to S_A and L_A represents the restriction of a description of what is globally the case in the bipartite system to a description of what is locally the case in Alice’s subsystem—this is a tool crucial for the purpose of this paper, of capturing non-locality and contextuality. We will see, for instance, that this projection has a role in characterizing the no-signalling property in fibrational terms in section 3.

It may need stressing that S_{AB} described above is just a cartesian product (taken fiberwise over L_\emptyset)—rather than anything similar to a tensor product—of S_A and S_B ; hence it does not by itself express any

correlation between Alice's and Bob's measurement outcomes. It is rather a transition-system expression for the 4×4 entries in a probability table describing a $(2, 2, 2)$ -scenario. Any correlation will be expressed by assigning probabilities to transitions in S_{AB} ; we will see how in section 3.

3 Adding Probabilities to Presheaves

This short section lays out how to add probabilities to the presheaf representation of non-deterministic processes given in section 2. The definitions provided here will later be generalized in subsection 4.2, after a generalization of the presheaf representation is proposed in subsection 4.1.

3.1 Stochastic Presheaves

Recall that in a description of a non-deterministic process with a presheaf $S : L^{\text{op}} \rightarrow \mathbf{Sets}$, for any edge $e = (x, y)$ of L and state $s \in S(x)$, the inverse image $S(e)^{-1}(s) \subseteq S(y)$ is the set of states to which the system may internally choose to transition from s when e is externally chosen. Now we want to give probability to such an internal choice; so let us achieve just that, with the following series of definitions. They use the notion of R -distribution for a commutative semiring R ; see [1, §2.3] for its definition. In particular, throughout this paper all distributions are assumed to be normalized and with finite support.

Definition 2. Fix a commutative semiring R . Given any sets X and Y , we define an R -map from X to Y as any surjection $f : Y \rightarrow X$ (note the opposite direction) equipped with, for each $s \in X$, an R -distribution on $f^{-1}(s) \subseteq Y$, written d_s^f . (We say that an R -map is *on* its underlying surjection.)

Obviously, we can achieve what we wanted above with an $\mathbb{R}_{\geq 0}$ -map f on $S(e) : S(y) \rightarrow S(x)$ (assuming $S(e)$ is surjective): The distribution d_s^f assigns to each $t \in S(e)^{-1}(s)$ the probability $d_s^f(t)$ with which the system transitions from s to t (when e is chosen). To do this for the entire presheaf, we give

Definition 3. Given any two R -maps f^R on $f : Z \rightarrow Y$ and g^R on $g : Y \rightarrow X$, let their composition $f^R \circ g^R$ be on $g \circ f : Z \rightarrow X$ with, for each $s \in X$, an R -distribution $d_s^{g \circ f}$ on $f^{-1}(g^{-1}(s)) \subseteq Z$ such that

$$d_s^{g \circ f}(u) = d_s^g(f(u)) \cdot d_{f(u)}^f(u). \quad (3)$$

Write **R -Map** for the category of sets and R -maps. (Clearly, the unique R -map on the identity map $1_X : X \rightarrow X$ is the identity on X in **R -Map**.)

The point of (3) should be clear: When $s = g(t)$ and $t = f(u)$, the system transitions from s to t with probability $r = d_s^g(t)$ and from t to u with probability $r' = d_t^f(u)$; so it transitions from s to u with probability $r \cdot r' = d_s^g(t) \cdot d_t^f(u) = d_s^{g \circ f}(u)$. (Note that the system can go from s to u through at most one t , since f is a function.) Then, finally,

Definition 4. An R -presheaf over a category \mathbf{C} is a contravariant functor from \mathbf{C} to **R -Map**. (We say that an R -presheaf is *on* its underlying presheaf.)

So, given a presheaf $S : L^{\text{op}} \rightarrow \mathbf{Sets}$ over a tree L as an L -LTS, we assign probabilities to the internal choices in S by simply taking an R -presheaf on S .

The presheaf-fibration equivalence (Fact 1) partially extends to R -presheaves: We can define “ R -fibrations” and prove that the equivalence extends to an essentially surjective and full functor from the category of rooted R -presheaves over a rooted tree L to that of rooted R -fibrations over L (we however omit the definitions and proof in this abstract). This extended version is limited and no longer an equivalence, but good enough for practical purposes. The core idea is that, given an R -presheaf $S : L^{\text{op}} \rightarrow \mathbf{Rel}$

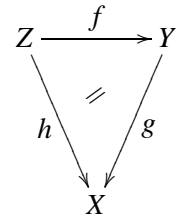
that has a root $s \in S(x)$, the “horizontal” assignment of probabilities $d_s^{S(x,y)}(t)$ to all states $t \in S$ can be turned into a “vertical” assignment of probabilities $d_y^\pi(t)$ on the fibration $\pi : S \rightarrow L$ that corresponds to the underlying presheaf of S .

Lastly, note that, although it may be proper to reserve the term “probability” to values of $\mathbb{R}_{\geq 0}$ -distributions, in this paper I apply the term broadly to values of R -distributions in general. Other interesting cases of R include \mathbb{B} , the booleans, and \mathbb{R} , all the reals, both of which are discussed in [1].

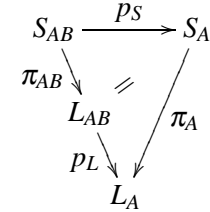
3.2 Example: No-Signalling

Let us say that a commutative semiring R is “normalizable” if, for every family $\{r_i\}_{i \in I}$ of elements of R such that $c := \sum_{i \in I} r_i \neq 0$, there is a family $\{r'_i\}_{i \in I}$ of elements of R such that $r'_j \cdot c = r_j$ for each $j \in I$ and $\sum_{i \in I} r'_i = 1$. For instance, $\mathbb{R}_{\geq 0}$ is normalizable. Now, in $R\text{-Map}$ for normalizable R , we have the following fact (a proof is omitted since it is straightforward).

Fact 2. Suppose R is normalizable. Then a factorization of a surjection into surjections, $h = g \circ f : Z \twoheadrightarrow Y \twoheadrightarrow X$, induces the following function ϕ : For any R -map h^R on h , $\phi(h^R)$ is the (unique) R -map on g through which h^R factors (in $R\text{-Map}$); that is, $h^R = f^R \circ \phi(h^R)$ for some R -map f^R on f . More explicitly, $\phi(h^R)$ is defined by $d_s^{\phi(h^R)}(t) = \sum_{u \in f^{-1}(t)} d_s^{h^R}(u)$ for $s \in X$ and $t \in Y$; in other words, $\phi(h^R)$ is the *marginal* of h^R along the identification of states $u \in Z$ by the quotient map $f : Z \twoheadrightarrow Y$. In addition, ϕ is a surjection from the R -maps on h to those on g .



Let us apply this fact to the diagram in (2), writing $p_S : S_{AB} \twoheadrightarrow S_A$ and $p_L : L_{AB} \twoheadrightarrow L_A$ for the pair of projections. Take $h = p_L \circ \pi_{AB}$ and $g = \pi_A$, with $f = p_S$. Then $\phi(h^R)$ (on π_A) is the marginal of h^R (on $p_L \circ \pi_{AB}$) along the restriction of description from the bipartite system to Alice’s. Note that, however, this involves probabilities on p_L , that is, with which Bob chooses from measurements b and b' . Different probabilities on p_L may lead to different $\phi(h^R)$ —or maybe not, if the probabilities on π_{AB} satisfy the no-signalling property. More precisely, we have the following (in which we write π_{AB} so as to connect to (2), but the system can consist of any number of parties).



Theorem 1. An R -presheaf π_{AB}^R on the presheaf π_{AB} for a multipartite system satisfies no-signalling iff, for each pair of projections p_S and p_L , $\phi(\pi_{AB}^R \circ p_L^R)$ is the same regardless of the choice of R -map p_L^R on p_L .

Proof. First observe that, for each $t \in S_A$, since $p_S^{-1}(t) = \sum_{v \in p_L^{-1}(\pi_A(t))} (p_S^{-1}(t) \cap \pi_{AB}^{-1}(v))$, we have

$$\begin{aligned}
 d_{\pi_A(t)}^{\phi(\pi_{AB}^R \circ p_L^R)}(t) &= \sum_{u \in p_S^{-1}(t)} d_{\pi_A(t)}^{\pi_{AB}^R \circ p_L^R}(u) = \sum_{u \in p_S^{-1}(t)} d_{\pi_A(t)}^{p_L^R}(\pi_{AB}(u)) \cdot d_{\pi_{AB}(u)}^{\pi_{AB}^R}(u) \\
 &= \sum_{v \in p_L^{-1}(\pi_A(t))} \sum_{u \in p_S^{-1}(t) \cap \pi_{AB}^{-1}(v)} d_{\pi_A(t)}^{p_L^R}(v) \cdot d_v^{\pi_{AB}^R}(u) \\
 &= \sum_{v \in p_L^{-1}(\pi_A(t))} d_{\pi_A(t)}^{p_L^R}(v) \cdot \sum_{u \in p_S^{-1}(t) \cap \pi_{AB}^{-1}(v)} d_v^{\pi_{AB}^R}(u). \quad (4)
 \end{aligned}$$

Now suppose π_{AB}^R satisfies no-signalling. This means that each $t \in S_A$ is assigned a real $e(t)$ such that every $v \in p_L^{-1}(\pi_A(t))$ satisfies $\sum_{u \in p_S^{-1}(t) \cap \pi_{AB}^{-1}(v)} d_v^{\pi_{AB}^R}(u) = e(t)$. (Note that this is independent of

choice of R -map on p_L .) So, plugging this into (4) we see that, for any R -map p_L^R on p_L , each $t \in S_A$ has $d_{\pi_A(t)}^{\phi(\pi_{AB}^R \circ p_L^R)}(t) = \sum_{v \in p_L^{-1}(\pi_A(t))} d_{\pi_A(t)}^{p_L^R}(v) \cdot e(t) = e(t)$, which is independent of the choice of p_L^R .

On the other hand, suppose $\phi(\pi_{AB}^R \circ p_L^R)$ is the same regardless of the choice of p_L^R . Fix any $t \in S_A$ and any $v_0, v_1 \in p_L^{-1}(\pi_A(t))$. Obviously, for each $i = 0, 1$, there is an R -map p_L^{Ri} on p_L such that $d_{\pi_A(t)}^{p_L^{Ri}}(v_i) = 1$. Then $\sum_{u \in p_S^{-1}(t) \cap \pi_{AB}^{-1}(v_0)} d_{v_0}^{\pi_{AB}^R}(u) = d_{\pi_A(t)}^{\phi(\pi_{AB}^R \circ p_L^{R0})}(t) = d_{\pi_A(t)}^{\phi(\pi_{AB}^R \circ p_L^{R1})}(t) = \sum_{u \in p_S^{-1}(t) \cap \pi_{AB}^{-1}(v_1)} d_{v_1}^{\pi_{AB}^R}(u)$ by (4). Since this holds for any pair of projections p_S and p_L , π_{AB}^R satisfies no-signalling. \square

4 Stochastic Relational Presheaves

In section 2 we saw how presheaves over trees—which are themselves trees—can be used as LTSs; and in section 3 we saw how to add probabilities to such systems. Generalizing this, this section obtains similarly labelled transition systems with probabilities that are however not trees.

4.1 Relational Presheaves

We first show how to implement LTSs of a non-tree shape using a presheaf-like structure. The core idea in using presheaves as LTSs was the following, functorial one: Let a tree L represent a series of external choices; assign to each stage in L the set of possible states at that stage; and connect states from different stages with internally chosen transitions. This idea involves no intrinsic reason why this connection of transitions should be (reverse) functional, i.e., why the functor we take should be a presheaf.

In fact, here is a reason the functor we take should *not* always be a presheaf. Consider the following two objectives, each of which may, conceivably, be well motivated.

- (i) For our functor S from the tree L , we may like to take, as values $S(x)$ for stages $x \in L$, the sets of states in Hilbert spaces instead of just any sets, to express quantum processes straightforwardly.
- (ii) We may consider a non-deterministic process that involves both branching and colliding (so cannot be a tree, forward or backward). In fact, when we do a quantum measurement a in one basis and then another a' in another basis, the system may transition from a state s to t_0 (after a) to u (after a'), but may also transition from s to $t_1 \neq t_0$ (after a) to the same u (after a').

The use of a presheaf, and in particular of functions $S(e)$ for edges e of L —which forces the transition system to be a tree—cannot accommodate both (i) and (ii). To accommodate a non-tree as in (ii) in a tree formalism, it is a standard technique to “unfold” or “unravel” the non-tree into a tree, duplicating the single state u to u_0 following t_0 and u_1 following t_1 . This, however, does not go well with (i), since the set $S(x)$ encompassing u_0, u_1 , and all the required duplicates may have to be much more complicated than just the set of states of a Hilbert space. This is why we should at least sometimes let $S(e)$, for edges e of L , be relations in general rather than functions. Then, in (ii), the state s can be connected to both t_0 and t_1 while both t_0 and t_1 connected to u .

So, instead of the category **Sets** of sets and functions, we take the category **Rel** of sets and relations as the codomain of our functors (see [6] and [10, esp. Ch. II] for categorical characterizations of **Rel** and its generalizations). For the sake of notation, let us enter

Definition 5. **Rel** is the category of sets and relations. Its objects are sets, and its arrows from a set X to another Y are relations $f \subseteq X \times Y$, written $f : X \rightarrowtail Y$ as well. We write $s \xrightarrow{f} t$ instead of $(s, t) \in f$, and, identifying $f : X \rightarrowtail Y$ with $f : X \rightarrow \mathcal{P}(Y)$, sometimes write $f(s) = \{t \in Y \mid s \xrightarrow{f} t\}$. The composition $g \circ f : X \rightarrowtail Z$ of $f : X \rightarrowtail Y$ and $g : Y \rightarrowtail Z$ is defined so that $s \xrightarrow{g \circ f} u$ iff $s \xrightarrow{f} t \xrightarrow{g} u$ for some $t \in Y$.

Rel is a dagger compact category. Firstly, it has a \dagger structure: Any $f : X \rightarrowtail Y$ has a unique opposite relation $f^\dagger : Y \rightarrowtail X$, so that $s \xrightarrow{f^\dagger} t$ iff $t \xrightarrow{f} s$. Also, even though the cartesian product is no longer the product in the categorical sense in **Rel**, it is still a monoidal product \otimes . In addition, the identification of $f : X \rightarrowtail Y$ with $f : X \rightarrow \mathcal{P}(Y)$ is just one aspect of the fact that **Rel** is the Kleisli category $\mathbf{Kl}(\mathcal{P})$ of the powerset monad \mathcal{P} on **Sets**. Now, let us finally provide

Definition 6. A *relational presheaf* over a category **C** is a covariant functor from **C** to **Rel**.¹

So we generalize presheaves with relational presheaves as our LTSs. We must note that relational presheaves are covariant and not contravariant. Thus, given an edge $e = (x, y)$ of a tree L , the system's transition from states at stage x to ones at stage y is represented by a relation $S(e) : S(x) \rightarrowtail S(y)$ in a relational presheaf $S : L \rightarrow \mathbf{Rel}$, whereas by a function $S(e) : S(y) \rightarrow S(x)$ in a presheaf $S : L^{\text{op}} \rightarrow \mathbf{Sets}$.

It may be worth noting that, although relational presheaves over a tree of labels are themselves LTSs, they are also a generalization of the ordinary kind of LTSs in the following sense. As Hermida [15] observes, given a set L of labels, the (ordinary) transition systems labelled by L are, in our terminology, the relational presheaves over the free monoid L^* generated by L . Our notion of relational presheaf as a LTS generalizes this by replacing L^* —a tree in which every (type of) edge is followed by every other (type of) edge—with a general tree, and permitting different stages to have different sets of states.

It is also worth noting that a small part of the presheaf-fibration equivalence (Fact 1) applies to relational presheaves, as relational presheaves over a tree L can be regarded as “open” bundles over L : The equivalence extends to an essentially surjective and faithful functor from the category of rooted and open bundles over a rooted tree L to that of rooted relational presheaves over L . (Again, we omit the definitions and proof in this abstract.)

4.2 Adding Probabilities to Relational Presheaves

We added probabilities to presheaves as LTSs in section 3. In this subsection, we add probabilities to relational presheaves, which we introduced in subsection 4.1. This can be done by simply replacing the functional elements of the definitions in section 3 with relational elements. (We should recall that, in the generalization given in subsection 4.1, a relation $f : X \rightarrowtail Y$ generalizes a function $f : Y \rightarrow X$ of the opposite direction.)

Definition 7. We define an *R-relation* from a set X to another Y as an “entire” relation $f : X \rightarrowtail Y$ (i.e., such that each $s \in X$ has some $t \in Y$ with $s \xrightarrow{f} t$; note that, unlike the case of R -maps, we do not flip the direction of f for R -relations) equipped with, for each $s \in X$, an R -distribution d_s^f on Y with support

$$\text{supp}(d_s^f) \subseteq f(s). \quad (5)$$

(We say that an R -relation is *on* its underlying relation.) Given two R -relations $f : X \rightarrowtail Y$ and $g : Y \rightarrowtail Z$, let their composition $g \circ f : X \rightarrowtail Z$ have, for each $s \in X$, an R -distribution $d_s^{g \circ f}$ on Z such that

$$d_s^{g \circ f}(u) = \sum_{t \in Y} d_s^f(t) \cdot d_t^g(u). \quad (6)$$

Write **R-Rel** for the category of sets and R -relations. (It should be clear that the unique R -relation on the identity relation $1_X : X \rightarrowtail X$ is the identity on X in **R-Rel**.)

¹Rosenthal [24] defines a relational presheaf as a “lax” functor; Sobociński [25] follows this “lax” definition in his account of relational presheaves as LTSs. In contrast, I define a relational presheaf “strongly”.

This notion of R -relation is closely related to that of *stochastic map*. We discuss this relationship in subsection 4.3; it will be significant to the discussion that (5) has “ \subseteq ” as opposed to “ $=$ ”.

Let us compare the equation (6) with the one (3) for R -maps. For R -maps $f : Z \rightarrow Y$ and $g : Y \rightarrow X$, there is at most one state $t \in Y$ through which the system may transition from a given $s \in X$ to a given $u \in Z$; so the probability of the transition from s to u is just the probability of this particular path, given by the product of the two transitions, from s to t and from t to u . In contrast, for R -relations $f : X \rightharpoonup Y$ and $g : Y \rightharpoonup Z$, there can be many paths through which the system may transition from $s \in X$ to $u \in Z$; yet, since these paths are mutually exclusive, we can just sum their probabilities up to obtain the probability of the transition from s to u . Lastly, enter

Definition 8. An R -relational presheaf over a category \mathbf{C} is a covariant functor from \mathbf{C} to $R\text{-}\mathbf{Rel}$. (We say that an R -relational presheaf is *on* its underlying relational presheaf.)

This definition provides a structure that integrates the three frameworks (i)–(iii) mentioned in Introduction: An R -relational presheaf $S : L \rightarrow R\text{-}\mathbf{Rel}$ over a tree L forms an L -LTS in which internal choices take place with probabilities and possibly in a non-tree fashion.

Example 1. Let a tree L represent a branching family of series of quantum measurements, gates, and other operations that can be performed. Then, for stages $x \in L$, let $S(x)$ be sets of states in (possibly, though not necessarily, identical) Hilbert spaces, and, for each edge $e = (x, y)$ of L , let $S(e) : S(x) \rightharpoonup S(y)$ be the $\mathbb{R}_{\geq 0}$ -relation that models the operation e in Hilbert-space terms, such as projections (branching with probabilities) to the suitable measurement basis. If L is moreover a free monoid and $S(x)$ are all identical (as in Hermida’s [15] formulation of transition systems mentioned in subsection 4.1), models amount essentially to ones given in Baltag and Smets [4].

This example gives a straightforward representation of quantum protocols. So it is not surprising at all that we can find non-local or contextual behaviors in such representations. Yet, using more general values than Hilbert spaces, R -relational presheaves can model not only the presence but also the absence of non-locality and contextuality, and indeed characterize contextuality, as we will see in section 5.

4.3 Relation to Other Work and Formulations

The notion of R -relation is closely related to that of *stochastic map*, or equivalently to Kleisli maps of the *distribution monad*.² A stochastic map from a set X to another Y is an X -indexed family of $\mathbb{R}_{\geq 0}$ -distributions on Y , with the composition defined exactly by (6). This can also be rewritten using

Definition 9. Given any set X , write $\mathcal{D}_R(X)$ for the set of R -distributions on X . This gives rise to the R -distribution functor $\mathcal{D}_R : \mathbf{Sets} \rightarrow \mathbf{Sets}$ (see [16] as well as [1, §2.3]), which is in fact a monad on \mathbf{Sets} (see [16]).

Then the stochastic maps f from a set X to another Y are exactly the functions $f : X \rightarrow \mathcal{D}_{\mathbb{R}_{\geq 0}}(Y)$, the Kleisli maps of $\mathcal{D}_{\mathbb{R}_{\geq 0}}$. Moreover the Kleisli composition amounts to (6), and so the category **Stoch** of sets and stochastic maps is the Kleisli category $\mathbf{Kl}(\mathcal{D}_{\mathbb{R}_{\geq 0}})$ of $\mathcal{D}_{\mathbb{R}_{\geq 0}}$ (see [17, §2]).

This is closely related to $R\text{-}\mathbf{Rel}$, but not exactly the same (aside from R generalizing $\mathbb{R}_{\geq 0}$): In short, an $\mathbb{R}_{\geq 0}$ -relation on a relation f is a stochastic map with an extra piece of information, namely, the underlying relation f . To express this formally, consider the following subfunctor of $\mathcal{P} \times \mathcal{D}_R : \mathbf{Sets} \rightarrow \mathbf{Sets}$.

$$T : X \mapsto \sum_{S \in \mathcal{P}(X)} \mathcal{D}_R(S) = \{ (S, d) \in \mathcal{P}(X) \times \mathcal{D}_R(X) \mid d \in \mathcal{D}_R(S) \}.$$

²I thank an anonymous referee for his/her comments regarding the relation between $R\text{-}\mathbf{Rel}$ and $\mathbf{Kl}(\mathcal{D}_{\mathbb{R}_{\geq 0}})$, which prompted me to write this subsection as a reply.

(We identify $d \in \mathcal{D}_R(X)$ and $d \in \mathcal{D}_R(S)$, as long as $\text{supp}(d) \subseteq S, X$.) Then the R -relations f from a set X to another Y are exactly the functions $f : X \rightarrow T(Y)$, with a $\mathcal{P}(Y)$ component. The two sets $T(Y)$ and $\mathcal{D}_R(Y)$ are related by the projection $p : (S, d) \mapsto d$ and a section $s : d \mapsto (\text{supp}(d), d)$, but $s \circ p \neq 1$ since we have “ \subseteq ” as opposed to “ $=$ ” in (5). Thus an R -relation $f : X \rightarrow T(Y)$ carries properly more information, of the underlying relation, than a stochastic map $f : X \rightarrow \mathcal{D}_R(Y)$. More categorically put, postcomposing p and s with Kleisli maps gives a retraction and a section of categories so that

Fact 3. $\mathbf{Stoch} = \mathbf{Kl}(\mathcal{D}_{\mathbb{R}_{\geq 0}})$ is a retract of $\mathbb{R}_{\geq 0}\text{-}\mathbf{Rel}$, but the retraction is not faithful.

The extra piece of information may appear redundant, as long as we are concerned with probabilities of transitions; yet that piece of information sometimes proves useful. In such a model as in (1) or (2), the underlying relational presheaf S describes the “logical” constraint of which states can be “logically” connected to which states; for instance, on the left of (1), state 00 can follow 0 but cannot 1. When we add the “physical” information of probabilities to S by taking an R -relational presheaf on S , the “logical” information is sometimes entailed by supports, but not always so: If the edge connecting states 0 and 00 in (1) has probability 0, then the support cannot tell us whether state 00 can “logically” follow state 0 or 1. It is useful to retain the “logical” constraint so as to consider a family of physical models satisfying it, as opposed to just one model—it is as useful as having a table of 4×4 entries that accommodates a family of probability assignments to outcomes in a $(2, 2, 2)$ -scenario. And for this purpose we need to retain the underlying relations, hence using $R\text{-}\mathbf{Rel}$ as opposed to \mathbf{Stoch} .

Lastly, it may be useful to note that the unit and multiplication of the cartesian product monad $\mathcal{P} \times \mathcal{D}_R$ restrict to the subfunctor T ,³ and that the composition in $R\text{-}\mathbf{Rel}$ is the Kleisli composition of T ; thus

Fact 4. T is a monad on \mathbf{Sets} , and $R\text{-}\mathbf{Rel}$ is its Kleisli category $\mathbf{Kl}(T)$.

This puts $R\text{-}\mathbf{Rel}$ in the tradition [18, 12, 21, 22, 8, 17, etc.] of using algebras for monads to represent stochastic relations.

5 Dynamic Logic for Contextuality

So far we have laid out R -presheaves and R -relational presheaves as labelled and stochastic transition systems. Now we demonstrate that these models are good enough for representing essential features of stochastic dynamics such as shown in quantum systems, by showing that they can characterize non-locality and contextuality; in fact, the dynamic logic of those transition systems is expressive enough to express this characterization in logical terms.

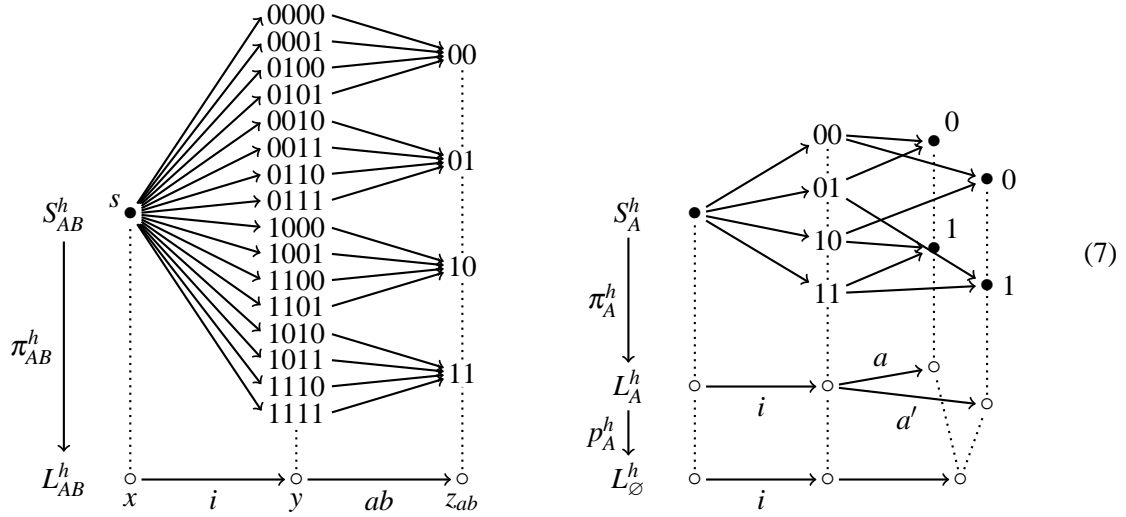
5.1 Deterministic Hidden-Variable Models

In their sheaf-theoretic approach to non-locality and contextuality, Abramsky and Brandenburger [1] provided a characterization of non-locality and contextuality in terms of “global sections” of certain presheaves; see [1, esp. §3 and §8]. We can “translate” this characterization into our setting of stochastic relational presheaves as follows.

Suppose we have an $\mathbb{R}_{\geq 0}$ -presheaf representing an “empirical model” for a (n, k, ℓ) -scenario that satisfies no-signalling in the sense of subsection 3.2. (The characterization given in [1] is more general than just about (n, k, ℓ) -scenarios, though I only take (n, k, ℓ) -scenarios here. We can translate the characterization in full generality, but omit it in this abstract.) As an example, let us take an $\mathbb{R}_{\geq 0}$ -presheaf E on the presheaf S_{AB} in (2) (and assume no-signalling). Then E is realized by a (factorizable) hidden-variable

³See Definition 2.1.2 of [19] for a concrete description of a cartesian product monad and its unit and multiplication.

model if and only if it has a “global section” (Theorem 8.1 of [1])—meaning, in our terms, that there exists an $\mathbb{R}_{\geq 0}$ -relational presheaf H on the relational presheaf S_{AB}^h in



(complete the picture by adding edges ab' , $a'b$, and $a'b'$ to L_{AB}^h) from which E is obtained by forgetting the middle stage y with the change-of-base operation as on the left of (1), that is, $E = H \circ m^{\text{op}}$ for the embedding $m : L_{AB} \hookrightarrow L_{AB}^h$ that omits y . Here $E = H \circ m^{\text{op}}$ means that $E(ab) = H(ab \circ i) = H(ab) \circ H(i)$, and hence that, by (6),

$$d_s^{E(ab)}(u) = \sum_{t \in H(y)} d_s^{H(i)}(t) \cdot d_t^{H(ab)}(u). \quad (8)$$

This is exactly to “reproduce the empirically observed probabilities $[d_s^{E(ab)}]$ by *averaging* over the hidden variables with respect to the distribution $[d_s^{H(i)}]$ ” ([1], p. 11).

From this characterization, the following features of H should be obvious: The set $H(y)$, which is forgotten in E , is a set of latent “instruction sets” (see [20]); moreover, they are *deterministic*, as $H(ab)$ is an $\mathbb{R}_{\geq 0}$ -relation on a *function*, as opposed to just any relation, from $H(y)$ to $H(z_{ab}) = E(z_{ab})$. Thus, the contextuality in a labelled and stochastic transition system E amounts to the failure of E to have such a deterministic hidden-variable model H . A little more formally,

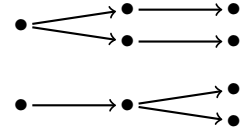
Theorem 2. *For an empirical model E (in the sense of [1]), the following are equivalent.*

- (i) *E has a realization by a factorizable hidden-variable model.*
- (ii) *E has a global section.*
- (iii) *The $\mathbb{R}_{\geq 0}$ -relational presheaf for E is obtained by forgetting the middle stage of a deterministic hidden-variable model.*

Proof. “(i) iff (ii)” is Theorem 8.1 of [1]. “(ii) iff (iii)” is essentially due to the fact that the equation for “averaging over” in [1] (p. 11) is identical to (8). \square

Note that the underlying relational presheaf S_{AB}^h of H (or any general ones for (n, k, ℓ) -scenarios) is not provided *ad hoc*, but canonically obtained, in the manner of (2), as the fibered cartesian product $S_A^h \otimes_{L_{\emptyset}^h} S_B^h$ of the obvious hidden-variable models $S_A^h : L_A^h \rightarrow \mathbf{Rel}$ for Alice (as in (7) above) and for Bob.

To extract an essential idea from the discussion so far, contextuality means, in transitional terms, that a model is inconsistent with the first shape of branching to the right (in which the system internally chooses from latent instruction sets before external choices are made), but has to have the second shape (in which the system internally chooses outcomes when external choices are made). And the distinction between these two shapes is one of the things modal logic is good at. Thus we carry on to consider the modal logic of our labelled and stochastic transition systems.



5.2 Dynamic Logic of Stochastic Relational Presheaf Models

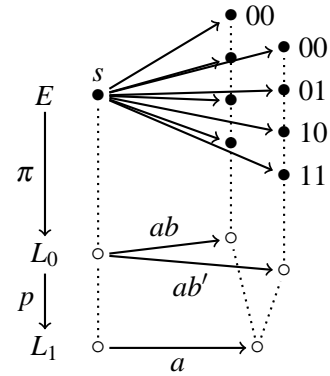
We lay out here how to use R -relational presheaves as a semantic structure for modal, dynamic logic. It turns out that the logic it gives rise to is expressive enough to capture in logical terms the characterization of contextuality we saw in subsection 5.1. (See [14] for general exposition of dynamic logic. A modal logic of stochastic relations expressed by algebras for a monad is also found in [8].)

Let us fix some (propositional) language; for our purposes it needs to have \wedge and \neg . Then we fix a set of labels e of transition (for instance, we use labels a, b, ab, ab' , etc., for a measurement scenario of $(2,2,2)$ -type). For each such label e , we add “dynamic modalities” $[e]$ and $\langle e \rangle$ to the language; we may also like to use probability modalities $P(- | e) \geq r$ for reals r . Since we take the base logic to be classical, $\langle e \rangle$ can be defined as $\neg[e]\neg$, and \top, \vee , biconditional \leftrightarrow and exclusive disjunction \oplus can be defined as usual. So we put

$$\varphi ::= p \mid \varphi \wedge \varphi \mid \neg\varphi \mid [e]\varphi \mid P(\varphi | e) > r \mid P(\varphi | e) = r \mid P(\varphi | e) < r$$

for propositional letters p and the labels e . For the sake of application to the contextuality in quantum measurements, we let each label be a measurement context (i.e., a jointly performable set of measurements), and each propositional letter have the form $a = k$ for a measurement a and an outcome k of a . (We will mention this application shortly in section 5; the semantics laid out in the remainder of this subsection can apply equally to other languages of the sort just defined.)

For this modal language, $\mathbb{R}_{\geq 0}$ -relational presheaves provide models. Firstly, the labels need interpreting in trees of labels. We take a family of trees—but not necessarily a single tree—and fibrations among them with an initial tree L_0 , so that each label e is an edge of one of the trees. For instance, the $(2,2,2)$ -scenario of Alice and Bob described in (2) has four trees L_- of labels and fibrations among them, with L_{AB} initial; the picture to the right describes $L_0 = L_{AB}$, $L_1 = L_A$, and the fibration $p : L_{AB} \rightarrow L_A$. Labels ab and ab' lie in L_0 , which takes both Alice and Bob as external to the system; a lies in L_1 , which takes Alice as external but Bob as internal. In general, p may fail to be a fibration or a function to L_1 (because, e.g., L_0 may have edges ab, bc, ca while L_1 has only a , so that bc cannot be projected down to any edge in L_1); but it has to be a partial function *onto* L_1 , so that p^\dagger is an entire relation on which there can be R -relations.



Then we take an R -relational presheaf $\pi : E \rightarrow L_0$ over the initial tree L_0 ; to keep describing the $(2,2,2)$ -scenario as an example, let us take E on S_{AB} in (2). Now, finally, we can provide interpretations $\llbracket \varphi \rrbracket$ for sentences φ of the language above by first assigning subsets $\llbracket p \rrbracket$ to propositional letters p and by then extending $\llbracket - \rrbracket$ recursively. We use the classical clauses for the Boolean connectives.

The new clauses of recursion for $\llbracket - \rrbracket$ concern $[e]\varphi$ and $P(\varphi | e) \geq r$. Since e may not lie in the initial tree L_0 , let $p : L_0 \rightarrow L_1$ be the (perhaps partial) function onto the tree L_1 in which e lies. Then we express

the ideas

- $[a] \varphi$ means that φ will be the case when Alice chooses a , regardless of which of b and b' Bob may choose;
- $P(\varphi \mid a) \gtrless r$ means that the probability with which φ will be the case when Alice chooses a is greater than (equal to, or less than) r , regardless of which of b and b' Bob may choose;

with the following clauses:

- $s \in \llbracket [e] \varphi \rrbracket$ iff $\text{supp}(d_s^{E(e')}) \subseteq \llbracket \varphi \rrbracket$ for all $e' = (\pi(s), x) \in p^{-1}(e)$;
- $s \in \llbracket P(\varphi \mid e) \gtrless r \rrbracket$ iff $\sum_{t \in S(e')(s) \cap \llbracket \varphi \rrbracket} d_s^{E(e')}(t) \gtrless r$ for all $e' = (\pi(s), x) \in p^{-1}(e)$.

It is worth noting that s may fail to be in any of $\llbracket P(\varphi \mid e) > r \rrbracket$, $\llbracket \dots = r \rrbracket$ and $\llbracket \dots < r \rrbracket$, when no-signalling fails (this is why the three sentences cannot define each other). On the other hand, $s \in \llbracket P(\varphi \mid a) = r \rrbracket$ implies that the model satisfies no-signalling regarding a .

Given this semantics, the following axioms and rules are sound (we omit ones regarding probability modalities; a complete axiomatization is an open problem):

- Classical propositional logic.
- Standard axioms and rules for every $[e]$:

$$\frac{\varphi \vdash \psi}{[e] \varphi \vdash [e] \psi}, \quad \vdash [e] \top, \quad [e] \varphi \wedge [e] \psi \vdash [e] (\varphi \wedge \psi).$$

- Moreover, whenever e_0, e_1 are such that $p(e_1) = e_0$ for one of the fibrations p ,

$$[e_0] \varphi \vdash [e_1] \varphi. \quad (9)$$

- In addition, because R -relations are on entire relations and distributions have nonempty supports, the semantics validates

$$\langle e_0 \rangle \langle e_1 \rangle \top \vdash [e_0] \langle e_1 \rangle \top.$$

5.3 Dynamic-Logical Characterization of Contextuality

The dynamic logic and its semantics introduced in subsection 5.2 can provide characterization for contextuality, taking advantage of the kind of semantic structures we studied in subsection 5.1.

Recall the characterization of contextuality in Theorem 2. That is, a model E fails to be contextual iff obtained from a deterministic hidden-variable model H by forgetting the middle stage—that is, iff consistent with the possibility that, at the middle stage, i.e., after i and before ab as in (7), the states are deterministic instruction sets. So, using labels and propositional letters for measurements and outcomes (as mentioned above in subsection 5.2), let us introduce the sentence Det expressing determinacy:

$$\text{Det}(a) := [a](a = 0) \vee [a](a = 1), \quad \text{Det} := \text{Det}(a) \wedge \dots \wedge \text{Det}(b').$$

Then the description of the Popescu-Rohrlich box [23] of $(2, 2, 2)$ -type,

$$\begin{aligned} \Delta_{\text{PR}} := & \langle ab \rangle \top \wedge \dots \wedge \langle a'b' \rangle \top \wedge [ab](a = 0 \leftrightarrow b = 0) \wedge [ab'](a = 0 \leftrightarrow b' = 0) \\ & \wedge [a'b](a' = 0 \leftrightarrow b = 0) \wedge [a'b'](a' = 0 \oplus b' = 0), \end{aligned}$$

entails $\neg \text{Det}$, using the axioms and rules mentioned above, including the suitable ones of the form (9) such as $[a] \varphi \vdash [ab] \varphi$. Also, a (partial) description Δ_{Hardy} of the Hardy model [13],

$$\begin{aligned} \Delta_{\text{Hardy}} := & \langle i \rangle \langle ab \rangle \top \wedge \cdots \wedge \langle i \rangle \langle a'b' \rangle \top \wedge \langle i \rangle \langle ab \rangle (a = 0 \wedge b = 0) \wedge [i] [ab'] (a = 1 \vee b' = 1) \\ & \wedge [i] [a'b] (a' = 1 \vee b = 1) \wedge [i] [a'b'] (a' = 0 \vee b' = 0) \end{aligned}$$

(note that this description involves label i), entails $\neg [i] \text{Det}$ using the same axioms.

We can generalize these examples. The upshot, roughly put, will be as follows.

- The sentence $\neg [i] \text{Det}$ characterizes contextuality.
- In addition, the sentence $[i] \neg \text{Det}$ characterizes “strong contextuality” (see [1, §6] for definition).

To put this more precisely and to prove it, we need some notation and definitions. First, fix a set M of measurements, along with a family of measurement contexts (i.e., jointly performable sets of measurements), and for each measurement a a set of outcomes O_a —we assume M and O_a to be finite. Then, in the vocabulary for M and O_a , write

- Λ for the set of axioms of the form (9), for any pair of measurement contexts $e_0, e_1 \subseteq M$ such that $e_0 \subseteq e_1$; and
- Det for the sentence $\bigwedge_{a \in M} \text{Det}(a)$, where $\text{Det}(a) := \bigvee_{k \in O_a} [a](a = k)$.

Moreover,

- By a “legal” sentence, let us mean a sentence of the form either $[i] [e] \varphi$, $\langle i \rangle \langle e \rangle \varphi$, or $P(\varphi \mid e \circ i) \geq r$ in which e is a jointly performable set of measurements and φ is a Boolean compound of propositional letters referring to no measurements other than those in e .

It is clear that legal sentences can be used to describe empirical models (again, in the sense of [1]). More precisely, let E be any empirical model involving $\mathbb{R}_{\geq 0}$ - (or \mathbb{B} -) distributions; i.e., for each measurement context e , E_e is a $\mathbb{R}_{\geq 0}$ - (or \mathbb{B} -) distribution on the set $\prod_{a \in e} O_a$. Then

- legal $[i] [e] \varphi$ describes E iff φ holds of every support of the distribution E_e ;
- legal $\langle i \rangle \langle e \rangle \varphi$ describes E iff φ holds of some support of the distribution E_e ;
- legal $P(\varphi \mid e \circ i) \geq r$ describes E iff $\sum_{f \in \text{supp}(E_e)} \varphi(f) E_e(f) \geq r$.

Then we finally have

Theorem 3. *Let Δ be a set of legal sentences that contains $\langle i \rangle \langle e \rangle \top$ for every maximal measurement context e . Then the following are equivalent.*

- Every empirical model E that Δ describes is contextual.
- Every stochastic relational presheaf model that satisfies no-signalling and validates Λ validates $\Delta \vdash \neg [i] \text{Det}$.

Moreover, the following are equivalent.

- Every empirical model E that Δ describes is strongly contextual.
- Every stochastic relational presheaf model that satisfies no-signalling and validates Λ validates $\Delta \vdash [i] \neg \text{Det}$.

Proof. Suppose (i) fails; that is, there is an empirical model E that Δ describes but that is not contextual, so that E has a global section E_M . Then construct a stochastic relational presheaf model $\pi : H \rightarrow L_0$ as follows. First build a tree L_0 with the edge $i : x \rightarrow y$ followed by the edges $e : y \rightarrow z_e$ for all the maximal measurement contexts e . Then, for each non-maximal measurement context e , build a tree L_e with i followed by $e : y \rightarrow z_e$. Between L_0 and L_e , we take a (typically partial) function $p_e : L_0 \rightarrow L_e$ that maps i to i and any e' such that $e \subseteq e'$ to e ; and, whenever $e_0 \subseteq e_1$, we take a function $p_{e_1, e_0} : L_{e_1} \rightarrow L_{e_0}$ that maps i to i and e_1 to e_0 . Now let $H(x) = \{s\}$; $H(y) = \prod_{a \in M} O_a$; $H(z_e) = \prod_{a \in e} O_a$; and, moreover, writing $|H(i)|$ and $|H(e)|$ for the underlying relations of the $\mathbb{R}_{\geq 0}$ - (or \mathbb{B} -) relations $H(i)$ and $H(e)$,

- Let $|H(i)|(s) = \prod_{a \in M} O_a$ and $d_s^{H(i)}(f) = E_M(f)$ for each $f \in \prod_{a \in M} O_a$.
- For each $f \in \prod_{a \in M} O_a$ and maximal context e , let $|H(e)|(f) = \{f \upharpoonright e\}$ (so that $d_f^{H(e)}$ is trivially deterministic).
- For each $a \in M$ and $k \in O_a$, let $\llbracket a = k \rrbracket = \sum_e$ is a maximal context and $a \in e \{g \in \prod_{a \in e} O_a \mid g(a) = k\}$.

Then it is straightforward to check that H satisfies no-signalling, that H validates Λ , and that $s \in \llbracket \psi \rrbracket$ for each $\psi \in \Delta$ since ψ describes E . Yet, since each $f \in \prod_{a \in M} O_a$ satisfies $f \in \bigcap_{a \in M} \llbracket [a](a = f(a)) \rrbracket \subseteq \llbracket \text{Det} \rrbracket$, we have $s \in \llbracket [i] \text{Det} \rrbracket$, so $s \notin \llbracket \neg [i] \text{Det} \rrbracket$. Therefore H does not validate $\Delta \vdash \neg [i] \text{Det}$. Thus (ii) fails.

On the other hand, assuming (ii) fails, let $\pi : H \rightarrow L_0$ be a stochastic relational presheaf model that validates Λ but that has some $s \in \bigcap_{\psi \in \Delta} \llbracket \psi \rrbracket$ with $s \notin \llbracket \neg [i] \text{Det} \rrbracket$. Since $s \in \llbracket [i] \langle e \rangle \top \rrbracket$ for each maximal context e , L_0 has edges $i : \pi(s) \rightarrow y$ and $e : y \rightarrow z_e$ for all the maximal contexts e . Then define a $\mathbb{R}_{\geq 0}$ - (or \mathbb{B} -) distribution E_M on $\prod_{a \in M} O_a$ so that, for each $f \in \prod_{a \in M} O_a$,

$$E_M(f) = \sum_{t \in |H(i)|(s) \cap \bigcap_{a \in M} \llbracket [a](a = f(a)) \rrbracket} d_s^{H(i)}(t).$$

Also, for each maximal context e , define a distribution E_e on $\prod_{a \in e} O_a$ so that, for each $f \in \prod_{a \in e} O_a$,

$$E_e(f) = \sum_{u \in |H(e)| \circ |H(i)|(s) \cap \bigcap_{a \in e} \llbracket [a](a = f(a)) \rrbracket} d_s^{H(e) \circ H(i)}(u).$$

Then it is easy to check that the family $E = \{E_e\}_e$ satisfies the compatibility condition (i.e., no-signalling) and hence is an empirical model, that Δ describes E , and that E_M is a global section for E . Thus (i) fails.

Suppose (iii) fails; that is, there is an empirical model E that Δ describes but that is not strongly contextual, so that there is a function $f : \prod_{a \in M} O_a$ such that $f \upharpoonright e \in \text{supp}(E_e)$ for every context e . Then construct a stochastic relational presheaf model $\pi : H \rightarrow L_0$ as follows. First build trees L_0 and L_e as in the first paragraph of this proof. Now let $H(x) = \{s\}$; $H(y) = \{f, t\}$; $H(z_e) = \prod_{a \in e} O_a$ for each maximal context e ; and, moreover, define $H(i)$ and each $H(e)$ as the following $\mathbb{R}_{\geq 0}$ - (or \mathbb{B} -) relations (we lay out how to define $\mathbb{R}_{\geq 0}$ -relations, since we can use their supports to define \mathbb{B} -relations).

- $|H(i)|(s) = \{f, t\}$, and $d_s^{H(i)}(f) = \min\{E_e(f \upharpoonright e) \mid e \text{ is a maximal context}\}$, so that $d_s^{H(i)}(t) = 1 - d_s^{H(i)}(f)$.
- $|H(e)|(f) = \{f \upharpoonright e\}$ (so that $d_f^{H(e)}$ is trivially deterministic).
- $|H(e)|(t) = \prod_{a \in e} O_a$. If $d_s^{H(i)}(t) = 0$, then $d_t^{H(e)}(g) = 1$ for all $g \in \prod_{a \in e} O_a$. Otherwise

$$d_t^{H(e)}(g) = \begin{cases} \frac{E_e(g) - d_s^{H(i)}(f)}{d_s^{H(i)}(t)} & \text{if } g = f \upharpoonright e, \\ \frac{E_e(g)}{d_s^{H(i)}(t)} & \text{otherwise.} \end{cases}$$

Lastly, for each $a \in M$ and $k \in O_a$, let $\llbracket a = k \rrbracket = \sum_e$ is a maximal context and $a \in e \{g \in \prod_{a \in e} O_a \mid g(a) = k\}$. Then it is straightforward to check that H satisfies no-signalling, that H validates Λ , and that $s \in \llbracket \psi \rrbracket$ for each $\psi \in \Delta$ since ψ describes E . Yet $f \in \bigcap_{a \in M} \llbracket [a](a = f(a)) \rrbracket \subseteq \llbracket \text{Det} \rrbracket$ implies $s \notin \llbracket [i] \neg \text{Det} \rrbracket$. Therefore H does not validate $\Delta \vdash [i] \neg \text{Det}$. Thus (iv) fails.

On the other hand, assuming (iv) fails, let $\pi : H \rightarrow L_0$ be a stochastic relational presheaf model that validates Λ but that has some $s \in \bigcap_{\psi \in \Delta} \llbracket \psi \rrbracket$ with $s \notin \llbracket [i] \neg \text{Det} \rrbracket$. As before, since $s \in \llbracket \langle i \rangle \langle e \rangle \top \rrbracket$ for each maximal context e , L_0 has edges $i : \pi(s) \rightarrow y$ and $e : y \rightarrow z_e$ for all the maximal contexts e . Then, as before, for each maximal context e , define a distribution E_e on $\prod_{a \in e} O_a$ so that, for each $f \in \prod_{a \in e} O_a$,

$$E_e(f) = \sum_{u \in |H(e)| \circ |H(i)|(s) \cap \bigcap_{a \in e} \llbracket a = f(a) \rrbracket} d_s^{H(e) \circ H(i)}(u).$$

Again it is easy to check that the family $E = \{E_e\}_e$ is an empirical model that Δ describes. Now note that, since $s \notin \llbracket [i] \neg \text{Det} \rrbracket$, that is, since $s \in \llbracket \langle i \rangle \text{Det} \rrbracket$, some $t \in \text{supp}(d_s^{H(i)})$ lies in $\llbracket \text{Det} \rrbracket$ and therefore we have a function $f \in \prod_{a \in M} O_a$ with which each $a \in M$ has $t \in \llbracket [a](a = f(a)) \rrbracket$; this implies that, for each maximal context e , there is $u \in |H(e)| \circ |H(i)|(s) \cap \bigcap_{a \in e} \llbracket a = f(a) \rrbracket$ such that $d_s^{H(e) \circ H(i)}(u) \geq d_s^{H(i)}(t) \cdot d_t^{H(e)}(u) > 0$, which means that $f|_e \in \text{supp}(E_e)$. Hence E is not strongly contextual. Thus (iii) fails. \square

6 Conclusion

In this paper, we have integrated the three frameworks mentioned in the Introduction for capturing non-deterministic processes, (i)–(iii), by introducing the category **R-Rel** of R -relations and taking R -relational presheaves—functors from trees to **R-Rel**. The resulting structure captures stochastic dynamics with a good enough expressive power, as demonstrated by the fact that it provides a labelled transitional formulation for the sheaf-theoretic approach of Abramsky and Brandenburger [1] to non-locality and contextuality, and moreover yielding dynamic logic with a modal-logical characterization of contextuality. (In fact, our formalism is partially equivalent to the sheaf-theoretic approach, extending the equivalence between presheaves and fibrations.) Whereas the sheaf-theoretic approach can take advantage of methods of cohomology to calculate conditions for contextuality (see [2]), our approach on the other hand has a certain flexibility in the base trees of measurement labels, so that it can readily express contextuality in not just one round of measurements but within a sequence or protocol of measurements. Thus our approach is expected to complement the sheaf-theoretic approach and extend it to various applications. Needless to say, applications to other kinds of stochastic dynamics can be expected as well.

References

- [1] Samson Abramsky & Adam Brandenburger (2011): *The Sheaf-Theoretic Structure of Non-Localizability and Contextuality*. *New Journal of Physics* 13:113036, doi:10.1088/1367-2630/13/11/113036.
- [2] Samson Abramsky, Shane Mansfield & Rui Soares Barbosa (2012): *Presheaf Models for Concurrency*. In Bart Jacobs, Peter Selinger & Bas Spitters, editors: *Proceedings 8th International Workshop on Quantum Physics and Logic (QPL 2011)*, *Electronic Proceedings in Theoretical Computer Science* 95, pp. 1–14, doi:10.4204/EPTCS.95.1.
- [3] John Baez, Tobias Fritz & Tom Leinster (2011): *A Characterization of Entropy in Terms of Information Loss*. Available at <http://arxiv.org/abs/1106.1791>.
- [4] Alexandru Baltag & Sonja Smets (2006): *LQP: The Dynamic Logic of Quantum Information*. *Mathematical Structures in Computer Science* 16, pp. 491–525, doi:10.1017/S0960129506005299.

- [5] Jonathan Barrett (2007): *Information Processing in Generalized Probabilistic Theories*. *Physical Review A* 75:032304, doi:10.1103/PhysRevA.75.032304.
- [6] Aurelio Carboni & Robert F. C. Walters (1987): *Cartesian Bicategories I*. *Journal of Pure and Applied Algebra* 49, pp. 11–32, doi:10.1016/0022-4049(87)90121-6.
- [7] Gian Luca Cattani & Glynn Winskel (1997): *Presheaf Models for Concurrency*. In Dirk van Dalen & Marc Bezem, editors: *Computer Science Logic: 10th International Workshop, CSL'96, Lecture Notes in Computer Science* 1258, Springer, pp. 58–75, doi:10.1007/3-540-63172-0_32.
- [8] Ernst-Erich Doberkat (2007): *Stochastic Relations: Foundations for Markov Transition Systems*. Chapman & Hall/CRC, doi:10.1201/9781584889427.
- [9] Brendan Fong (2012): *Causal Theories: A Categorical Perspective on Bayesian Networks*. Master's thesis, University of Oxford.
- [10] Peter J. Freyd & Andre Scedrov (1990): *Categories, Allegories*. North-Holland.
- [11] Tobias Fritz (2009): *A Presentation of the Category of Stochastic Matrices*. Available at <http://arxiv.org/abs/0902.2554>.
- [12] Michèle Giry (1982): *A Categorical Approach to Probability Theory*. In B. Banaschewski, editor: *Categorical Aspects of Topology and Analysis, Lecture Notes in Mathematics* 915, Springer, pp. 68–85, doi:10.1007/BFb0092872.
- [13] Lucien Hardy (1993): *Nonlocality for Two Particles without Inequalities for Almost All Entangled States*. *Physical Review Letters* 71, pp. 1665–1668, doi:10.1103/PhysRevLett.71.1665.
- [14] David Harel, Dexter Kozen & Jerzy Tiuryn (2000): *Dynamic Logic*. MIT Press.
- [15] Claudio Hermida (2011): *A Categorical Outlook on Relational Modalities and Simulations*. *Information and Computation* 209, pp. 1505–1517, doi:10.1016/j.ic.2010.09.009.
- [16] Bart Jacobs (2010): *Convexity, Duality and Effects*. In Christian S. Calude & Vladimiro Sassone, editors: *Theoretical Computer Science: 6th IFIP TC 1/WG 2.2 International Conference, TCS 2010*, Springer, pp. 1–19, doi:10.1007/978-3-642-15240-5_1.
- [17] Bart Jacobs (2011): *Probabilities, Distribution Monads, and Convex Categories*. *Theoretical Computer Science* 412, pp. 3323–3336, doi:10.1016/j.tcs.2011.04.005.
- [18] F. William Lawvere (1962): *The Category of Probabilistic Mappings*. Unpublished manuscript.
- [19] Ernie Manes & Philip Mulry (2007): *Monad Compositions I: General Constructions and Recursive Distributive Laws*. *Theory and Applications of Categories* 18, pp. 172–208. Available at <http://www.tac.mta.ca/tac/volumes/18/7/18-07abs.html>.
- [20] N. David Mermin (1981): *Quantum Mysteries for Anyone*. *Journal of Philosophy* 78, pp. 397–408, doi:10.2307/2026482.
- [21] Prakash Panangaden (1998): *Probabilistic Relations*. In C. Baier, M. Huth, M. Kwiatkowska & M. Ryan, editors: *Preliminary Proceedings of PROBMIV'98*, pp. 59–74.
- [22] Prakash Panangaden (1999): *The Category of Markov Kernels*. *Electronic Notes in Theoretical Computer Science* 22, pp. 171–187, doi:10.1016/S1571-0661(05)80602-4.
- [23] Sandu Popescu & Daniel Rohrlich (1994): *Quantum Nonlocality as an Axiom*. *Foundations of Physics* 24, pp. 397–385, doi:10.1007/BF02058098.
- [24] Kimmo I. Rosenthal (1996): *The Theory of Quantaloids*. Addison Wesley.
- [25] Paweł Sobociński (2012): *Relational Presheaves as Labelled Transition Systems*. In Dirk Pattinson & Lutz Schröder, editors: *Coalgebraic Methods in Computer Science, Lecture Notes in Computer Science* 7399, Springer, pp. 40–50, doi:10.1007/978-3-642-32784-1_3.
- [26] Daniele Varacca (2003): *Probability, Nondeterminism and Concurrency: Two Denotational Models for Probabilistic Computation*. Ph.D. thesis, University of Aarhus.

- [27] Glynn Winskel & Mogens Nielsen (1997): *Presheaves as Transition Systems*. In Doron Peled, Vaughan R. Pratt & Gerard J. Holzmann, editors: *Partial Order Methods in Verification: DIMACS Workshop July 24–26, 1996*, American Mathematical Society, pp. 129–140.